Computational methods for comparing and integrating multiple probing assays to predict RNA secondary structure

Afaf Saaidi[1,2,3], Delphine Allouche[3,4]
Supervisors: Yann Ponty [1,2,3], Bruno sargueil[3,4], Mireille Regnier[1,2]
1-AMIB team, Inria Saclay **2**-LIX ,**3**-CNRS ,**4**-Laboratoire de cristallographie et RMN Biologiques, Paris Descartes

1-Introduction

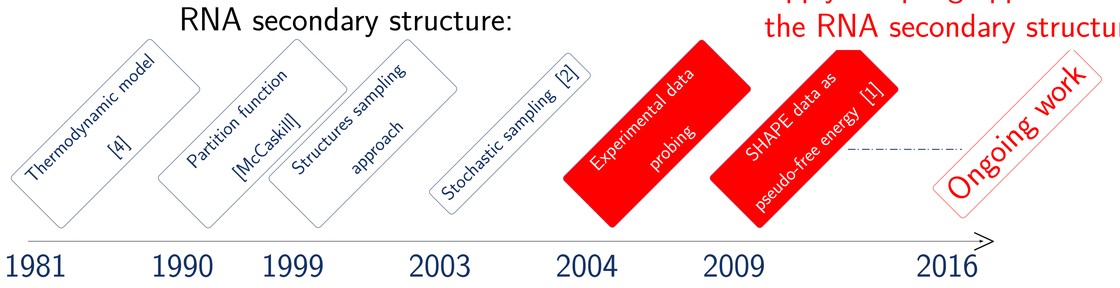
- RNA is key to understand many biological processes.
- RNA maintains a stable tertiary structure.
- The determination of the structure allows understanding its operating mechanism.
- We study the 444nt long VIH1 Gag-IRES.

RNA Structure determination

- 3D structure can be resolved experimentally [remains expensive and time-consuming].
- Computational methods allows to have accurate secondary structure predictions (PPV $\approx 75\%$). Less accurate predictions for long RNA.
- + Experimental Data [Chemical(SHAPE) \ Enzymatic] improve predictions.

State of the art

Evolution of computational approaches to predict the RNA secondary structure:



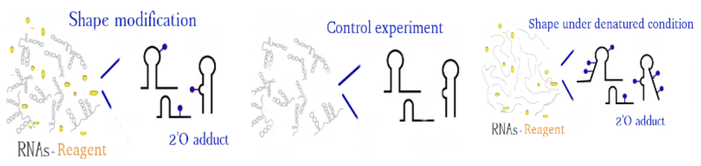
Objectives

- Apply sampling approach with SHAPE data to predict the RNA secondary structure.

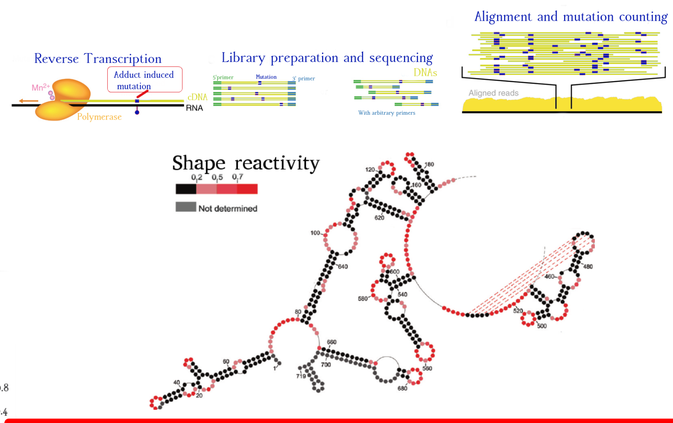
2-Material & Methods

2-1 Experimental data

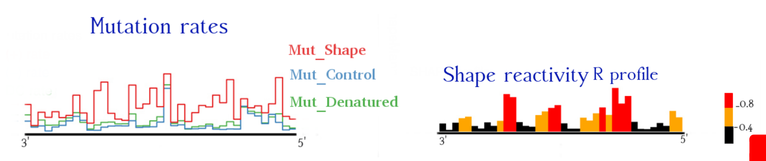
SHAPE-Map experiments



High Throughput Sequencing



SHAPE reactivity calculation

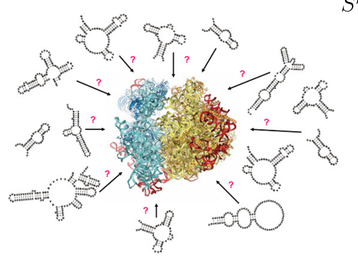


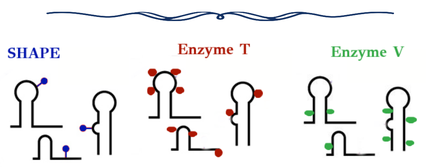
$$\textbf{Reactivity}(n) = \frac{mut_{\text{Shape}}(n) - mut_{\text{Control}}(n)}{mut_{\text{Denatured}}(n)} \quad [3]$$

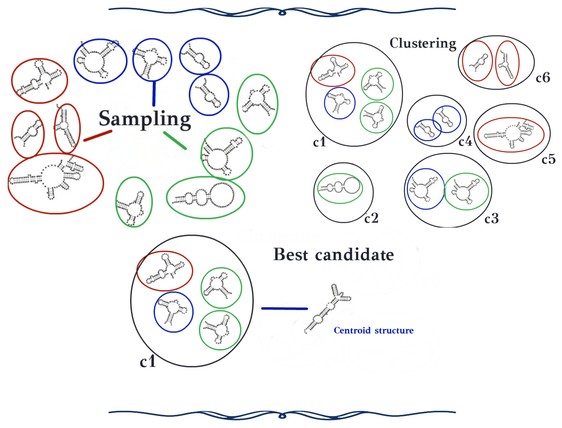
Boltzmann probability to observe a structure $\mathcal{S}$:

$$P(\mathcal{S}) = \frac{e^{\frac{-E_{\mathcal{S}}}{kT}}}{\mathcal{Z}}.$$

with $\mathcal{Z}$ the partition function: $\mathcal{Z} = \sum_{S' \in \mathcal{S}} e^{\frac{-E_{S'}}{kT}}$.





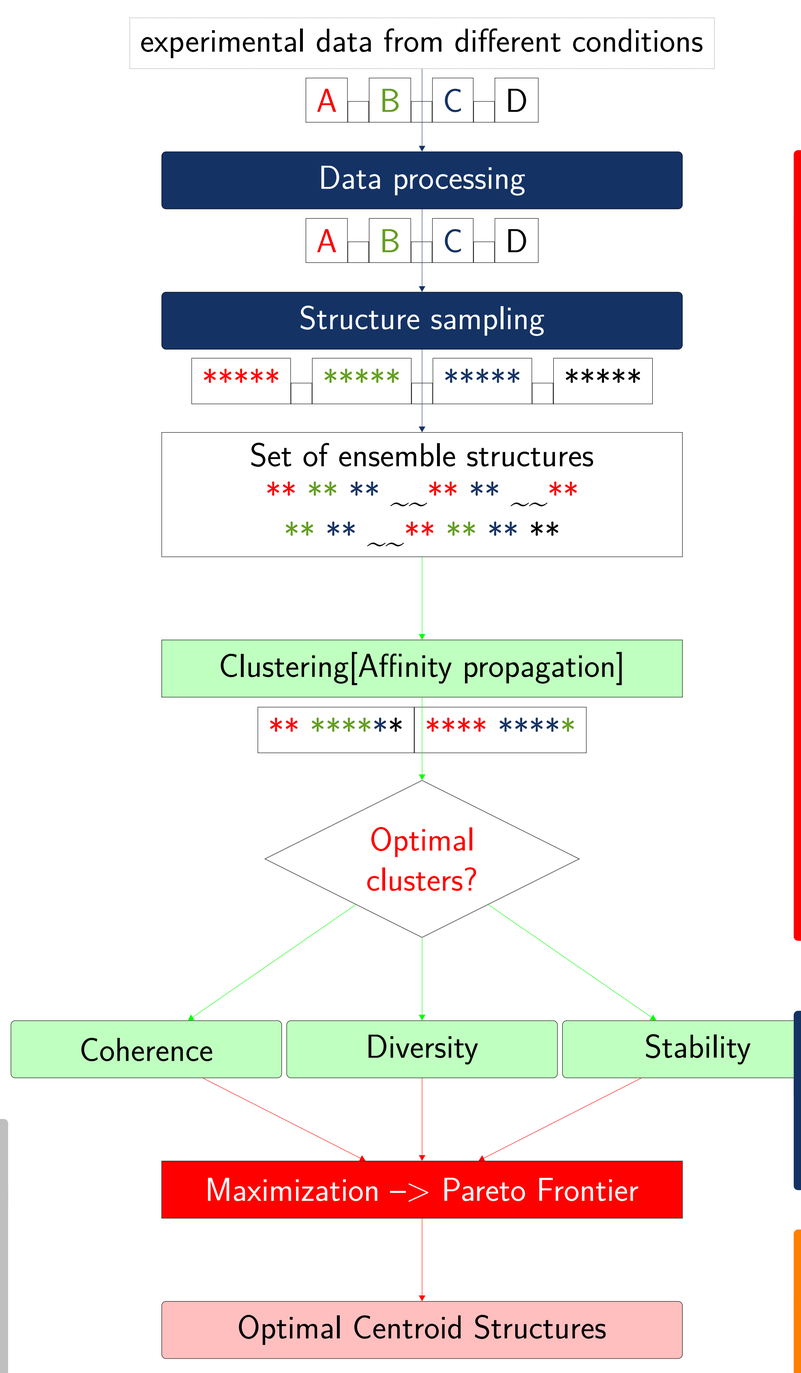


For a cluster $\mathcal{C}$

- Coherence ≡ Base pairs **Mean Distance** in $\mathcal{C}$.
- Diversity ≡ Number of **present conditions**.
- Stability ≡ Sum of **Boltzmann probabilities**.

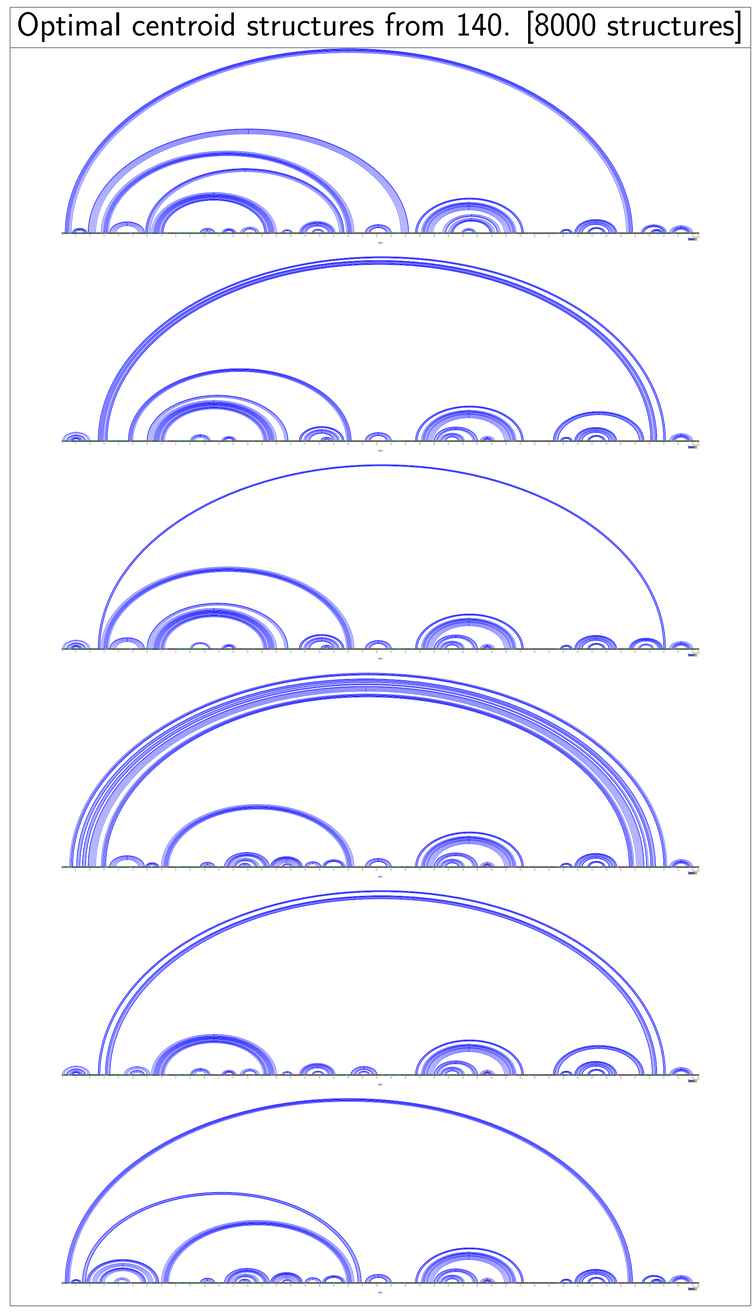
V-Enzymatic cleavage targets paired nucleotides.
T-Enzymatic cleavage reveals unpaired nucleotides.

2-2 Sampling/Clustering workflow



3-Results

Optimal centroid structures from 140. [8000 structures]



4-Conclusion & perspectives

- We have obtained a set of models supported by our integrative approach, those models are subject to validation.
- Some of centroid structures have shown high compatibility with existing proposed structural models.
- We will extend the approach to the simultaneous analysis of probing data for a set of RNA variants.

References

[1] Katherine E. Deigan, Tian W. Li, David H. Mathews, and Kevin M. Weeks.
Accurate SHAPE-directed RNA structure determination.
106(1):97–102, jan 2009.

[2] Lawrence CE. Ding.
A statistical sampling algorithm for RNA secondary structure prediction.
Nucleic Acids Research, 31(24):7280–7301, 2003.

[3] Matthew J Smola, Greggory M Rice, Steven Busan, Nathan A Siegfried, and Kevin M Weeks.
Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis.
Nature protocols, 10(11):1643–1669, 2015.

[4] Michael Zuker and Patrick Stiegler.
Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information.
Nucleic Acids Research, 9(1):133–148, jan 1981.

Acknowledgments

PhD funded by the "Fondation pour la Recherche Médicale".

Contact address

saaidi.afaf@polytechnique.edu